



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2016

Defining Benchmarks for Major Liver Surgery

Rössler, Fabian ; Sapisochin, Gonzalo ; Song, GiWon ; Lin, Yu-Hung ; Simpson, Mary Ann ; Hasegawa, Kiyoshi ; Laurenzi, Andrea ; Sánchez Cabús, Santiago ; Nunez, Milton Inostroza ; Gatti, Andrea ; Beltrame, Magali Chahdi ; Slankamenac, Ksenija ; Greig, Paul D ; Lee, Sung-Gyu ; Chen, Chao-Long ; Grant, David R ; Pomfret, Elizabeth A ; Kokudo, Norihiro ; Cherqui, Daniel ; Olthoff, Kim M ; Shaked, Abraham ; García-Valdecasas, Juan Carlos ; Lerut, Jan ; Troisi, Roberto I ; De Santibanes, Martin ; Petrowsky, Henrik ; Puhan, Milo A ; Clavien, Pierre-Alain

DOI: <https://doi.org/10.1097/SLA.0000000000001849>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-125991>

Journal Article

Published Version

Originally published at:

Rössler, Fabian; Sapisochin, Gonzalo; Song, GiWon; Lin, Yu-Hung; Simpson, Mary Ann; Hasegawa, Kiyoshi; Laurenzi, Andrea; Sánchez Cabús, Santiago; Nunez, Milton Inostroza; Gatti, Andrea; Beltrame, Magali Chahdi; Slankamenac, Ksenija; Greig, Paul D; Lee, Sung-Gyu; Chen, Chao-Long; Grant, David R; Pomfret, Elizabeth A; Kokudo, Norihiro; Cherqui, Daniel; Olthoff, Kim M; Shaked, Abraham; García-Valdecasas, Juan Carlos; Lerut, Jan; Troisi, Roberto I; De Santibanes, Martin; Petrowsky, Henrik; Puhan, Milo A; Clavien, Pierre-Alain (2016). Defining Benchmarks for Major Liver Surgery. *Annals of Surgery*, 264(3):492-500.

DOI: <https://doi.org/10.1097/SLA.0000000000001849>

Defining Benchmarks for Major Liver Surgery

A multicenter Analysis of 5202 Living Liver Donors

Fabian Rössler, MD,* Gonzalo Sapisochin, MD,† GiWon Song, MD,‡ Yu-Hung Lin, MD,§ Mary Ann Simpson, MD, PhD,¶ Kiyoshi Hasegawa, MD, PhD,|| Andrea Laurenzi, MD,** Santiago Sánchez Cabús, MD, PhD,†† Milton Inostroza Nunez, MD,‡‡ Andrea Gatti, MD,§§ Magali Chahdi Beltrame, MD,¶¶ Ksenija Slankamenac, MD, PhD,* Paul D. Greig, MD,† Sung-Gyu Lee, MD, PhD,‡ Chao-Long Chen, MD, PhD,§ David R. Grant, MD,† Elizabeth A. Pomfret, MD, PhD,¶ Norihiro Kokudo, MD, PhD,|| Daniel Cherqui, MD,** Kim M. Olthoff, MD,|||| Abraham Shaked, MD,|||| Juan Carlos García-Valdecasas, MD, PhD,†† Jan Lerut, MD, PhD,‡‡ Roberto I. Troisi, MD, PhD,§§ Martin De Santibanes, MD,¶¶ Henrik Petrowsky, MD,* Milo A. Puhan, MD, PhD,*** and Pierre-Alain Clavien, MD, PhD*

Objective: To measure and define the best achievable outcome after major hepatectomy.

Background: No reference values are available on outcomes after major hepatectomies. Analysis in living liver donors, with safety as the highest priority, offers the opportunity to define outcome benchmarks as the best possible results.

Methods: Outcome analyses of 5202 hemi-hepatectomies from living donors (LDs) from 12 high-volume centers worldwide were performed for a 10-year period. Endpoints, calculated at discharge, 3 and 6 months postoperatively, included postoperative morbidity measured by the Clavien-Dindo classification, the Comprehensive Complication Index (CCI), and liver failure according to different definitions. Benchmark values were defined as the 75th percentile of median morbidity values to represent the best achievable results at 3 month postoperatively.

Results: Patients were young ($34 \pm [9]$ years), predominantly male (65%) and healthy. Surgery lasted $7 \pm [2]$ hours; 2% needed blood transfusions.

Mean hospital stay was $11.7 \pm [5]$ days. 12% of patients developed at least 1 complication, of which 3.8% were major events (\geq grade III, including 1 death), mostly related to biliary/bleeding events, and were twice higher after right hepatectomy. The incidence of postoperative liver failure was low. Within 3-month follow-up, benchmark values for overall complication were $\leq 31\%$, for minor/major complications $\leq 23\%$ and $\leq 9\%$, respectively, and a CCI ≤ 33 in LDs with complications. Centers having performed ≥ 100 hepatectomies had significantly lower rates for overall (10.2% vs 35.9%, $P < 0.001$) and major (3% vs 12.1%, $P < 0.001$) complications and overall CCI (2.1 vs 8.5, $P < 0.001$).

Conclusions: The thorough outcome analysis of healthy LDs may serve as a reference for evaluating surgical performance in patients undergoing major liver resection across centers and different patient populations. Further benchmark studies are needed to develop risk-adjusted comparisons of surgical outcomes.

Keywords: Comprehensive Complication Index, left hemi-hepatectomy, living donor, right hemi-hepatectomy

(*Ann Surg* 2016;264:492–500)

From the *Swiss HPB Center Zurich, Department of Surgery and Transplantation, University Hospital and University of Zurich, Zurich, Switzerland; †Department of Surgery, Toronto General Hospital, University of Toronto, Toronto, Ontario, Canada; ‡Department of Surgery, University of Ulsan College of Medicine, Seoul, Republic of Korea; §Department of Surgery, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan; ¶Department of Transplantation, Lahey Hospital and Medical Center, Burlington, MA; ||Artificial Organ and Transplantation Division and Hepato-Biliary-Pancreatic Surgery, University of Tokyo, Tokyo, Japan; **The Hepatobiliary Center, Department of Surgery, Paul Brousse Hospital, University Paris Sud, Villejuif, France; ††Department of Surgery, Hospital Clínic de Barcelona, Barcelona, Spain; ‡‡Department of Abdominal and Transplantation Surgery, University Hospitals Saint Luc, Brussels, Belgium; §§Department of General Hepatobiliary and Liver Transplantation Surgery, Ghent University Hospital and Medical School, Ghent, Belgium; ¶¶Department of Surgery, Division of HPB Surgery, Liver Transplant Unit, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina; ||||Department of Surgery, Division of Transplant Surgery, University of Pennsylvania, Philadelphia, PA; and ***Department of Epidemiology, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland.

Reprints: Pierre-Alain Clavien, MD, PhD, Departement of Surgery and Transplantation, University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland. E-mail: clavien@access.uzh.ch.

All the authors of this manuscript declare no conflicts of interest and no funding was received for this study.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0003-4932/14/26105-0821

DOI: 10.1097/SLA.0000000000001849

I ncreasing demands for quality assessment in surgery has led to the search for new methodologies to convincingly evaluate complex and expensive procedures. Comparison of results between centers, different periods of times or alternative therapies are currently unreliable. One central drawback is the lack of standard outcome measures or metrics, against which the results of a procedure may be convincingly compared. For example, although centers may report on large series of liver resections and claim superior outcome, the “optimal” or best achievable postoperative results remain speculative.

This leads to the concept of benchmarking, which is a process of measuring performance across health care providers, with the aim of improving performance and reduce unwarranted variation in health care delivery.¹ Often benchmarking is used to compare 1 endpoint—for example, 1-year survival—of a procedure for a specific center versus the average or median data of other centers in a region or a country.² Although informative—with a main aim to rouse competition and a search for improvement in individual centers—such an approach fails to present a wider spectrum of relevant endpoints as well as comparison with the best achievable results.

With this in mind, we embarked in a novel approach identifying relevant benchmark endpoints for complex and well-defined procedures. We propose living donor (LD) hepatectomy as a unique model for benchmarking in liver surgery. We initiated this work for

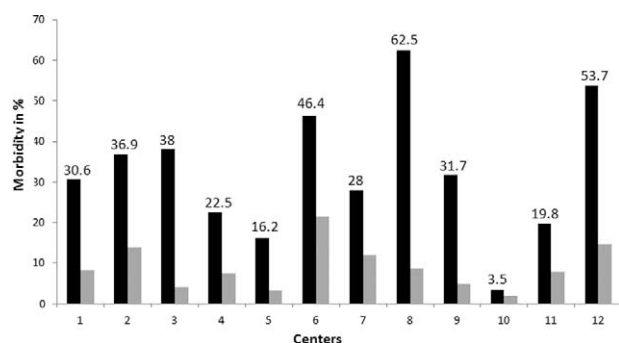


FIGURE 1. Shows the overall and major (\geq III) complication rate in % in the respective centers within 3 months follow up. Black indicates overall complication rate; Grey = major complication rate.

anatomic right hepatectomy (RH) and left hepatectomy (LH). Our hypothesis (starting point) was that the “best” achievable results come from LD operations as only standardized types of hepatectomies are performed, with safety as the highest priority. Such operations are conducted with maximum attention to all details and exclusively in healthy, low-risk individuals. Second, we selected only centers with high experience, having performed at least 50 cases of adult LD, and included centers covering 4 continents. Third, we identified a number of clinically relevant and reproducible endpoints. With these requirements, we postulated that the results of this complex procedure performed in such large population under the best available settings may serve to establish the optimal—better labeled “best achievable”—benchmark for major hepatectomies (Figs. 1 and 2).

METHODS

Study Design

Data were obtained in 12 high-volume liver centers covering the 4 continents for a 10-year period (January 1, 2004–December 31, 2014) with an experience of at least 50 LD operations. Each center had a prospectively collected database, from which most of the data was extracted. Charts were thoroughly reevaluated in each patient, who had a complicated postoperative course or unclear information. Only patients who underwent an anatomical RH or LH were analyzed. Pure laparoscopic procedures were excluded. The centers included the Hospital Clinic de Barcelona, Spain; Hospital Italiano de Buenos Aires, Argentina; University Hospitals Saint Luc Brussels, Belgium; University Hospital Ghent, Belgium; Lahey Hospital and Medical Center Burlington, USA; Kaohsiung Chang Gung Memorial Hospital,

Kaohsiung, Taiwan; Paul Brousse Hospital, Paris, France; University of Pennsylvania, USA; University of Tokyo, Japan; Toronto General Hospital, Canada; University of Ulsan Seoul, Korea; and the University Hospital of Zurich, Switzerland.

The aim of the study was to define specific benchmark values for well-defined endpoints describing postoperative morbidity after a major hepatectomy. If centers had higher postoperative morbidity than defined by the benchmark values, it would indicate a need to enquire about potential causes and quality improvement processes. If centers had lower postoperative morbidity than defined by the benchmark values, it would mean that postoperative morbidity is acceptable.

In agreement with each participating center, center-specific data are reported anonymously in this study, that is, without disclosure of the identity of the respective centers. Appropriate ethical approval from the institutional ethical board of the University Hospital of Zurich and from each respective center was obtained before analysis of the data.

Performance Metrics of Benchmarking

Five parameters of postoperative morbidity including complications of any severity, minor and major complications, the Comprehensive Complication Index (CCI), and postoperative liver failure served as performance metrics to define benchmarks for open RH and LH. All postoperative morbidity data were collected for the postoperative period of hospitalization and for the 3- and 6-month period after surgery.

Each postoperative complication occurring during the initial stay or after readmission was graded according to the severity-oriented Clavien-Dindo complication system.^{3,4} Minor and major complications were defined as \leq grade II and \geq grade III, respectively. Furthermore, the cumulative postoperative morbidity was assessed using the novel metric for postoperative complications—the CCI^{5,6}, which measures the overall morbidity on a scale from 0 (uneventful) to 100 (death).

LD characteristics including age, sex, comorbidities, preoperative laboratory tests, and volumetry of the right and left hemiliver, as well as available percentage of (macro and micro) steatosis obtained on biopsies were collected. Characteristics of surgery included following variables: duration of operation, transection technique and device, need for Pringle maneuver, intraoperative blood loss, and transfusion, as well as any “atypical” intraoperative events. Particular attention was given to the preservation of the middle hepatic vein for the donor or recipient. Postoperative laboratory values were collected for aspartate aminotransferase, alanine aminotransferase, bilirubin, alkaline phosphatase, gamma glutamyl transferase, creatinine, international normalized ratio (INR), hematocrit, hemoglobin, and platelet count.

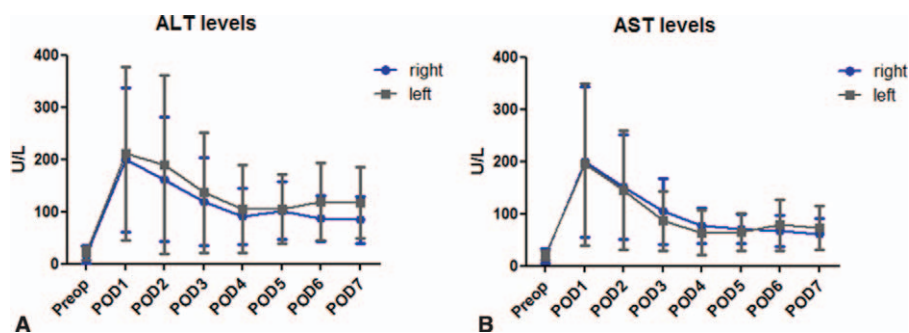


FIGURE 2. (A) and (B) show mean alanine aminotransferase and aspartate aminotransferase levels after right (RH) and left hemihepatectomy (LH) from preoperative to postoperative day 7.

TABLE 1. Patient Characteristics

	All (n = 5202)	RH (n = 4206)	LH (n = 996)
Age, yrs [SD]	30.9 [9.8]	30.0 [9.7]	34.1 [9.7]
Sex (male/female)	3189 (61.3%) / 2013 (38.7%)	2535 (60.3%) / 1671 (39.7%)	654 (65.7%) / 342 (34.3%)
ASA score			
≤ II	5195 (99.9%)	4200 (99.9%)	995 (99.9%)
≥ III	7 (0.1%)	6 (0.1%)	1 (0.1%)
BMI, kg/cm ² [SD]	23.5 [3.5]	23.4 [3.5]	24.4 [3.3]

Post-hepatectomy liver failure (PHLF) was assessed by 4 commonly reported criteria including International Study Group for Liver Surgery (ISGLS),⁷ 50–50 criteria,⁸ and bilirubin >7,⁹ as well as our criteria for small-for-size syndrome (SFSS) described in 2005,¹⁰ which are defined as following: ISGLS criteria are INR and bilirubin above the normal cutoff, as defined by local laboratory on day 5 after liver resection. Because our data relate to many different local laboratory standards, we set an INR ≥1.3 and bilirubin ≥1.2 mg/dL as thresholds,⁷ and the presence of biliary complications were not considered as an exclusion criteria. The 50–50 criteria are defined as INR ≥1.7 (Quick 50%) and serum bilirubin ≥2.9 mg/dL (50 μmol/L) on postoperative day 5, predicting in the original report a mortality of 50%.⁸ The bilirubin 7 criteria relies on serum bilirubin levels >7 mg/dL (119 μmol/L) on postoperative day 5 in noncirrhotic and non-cholestatic patients.⁹ Our definition of SFSS relies on 3 factors: bilirubin >5.9 mg/dL (100 μmol/L), INR >2 and the presence of encephalopathy grade 3 or 4. At least 2 of those factors must be present on 3 consecutive days within the first week after surgery.¹⁰

Statistical Analysis

Benchmark values were determined for the 5 parameters of postoperative morbidity, that is, complications of any severity, minor and major complications, the CCI and liver failure, all of them until hospital discharge and up to 3 and 6 months after surgery. For each center, we first calculated the proportion of LD with at least 1 complication of any severity, minor and major severity, liver failure as well as the median CCI of patients with at least 1 grade II complication. We then calculated the median proportions and the median of the median CCIs across centers and defined the benchmark range by the 25th and 75th percentile. Thus, morbidity values (eg, proportion of patients with a major complication) below the benchmark value (75th percentile) indicate acceptable morbidity, values

above the benchmark value indicate high morbidity, whereas below the benchmark value (25th percentile) indicate very low morbidity.

We also conducted descriptive statistics for various intra- and postoperative parameters in order to characterize the collective of LDs because it is the most comprehensive described so far.

RESULTS

From the 12 centers included in the study, we collected a series of 5202 LD cases, of which 4206 (81%) were anatomical RH and 996 (19%) anatomical LH. The numbers per center varied widely between 36 and 3257. Three centers performed more than 400 procedures, whereas 7 performed less than 100.

LD characteristics are summarized in Table 1. Of note, this population was at low surgical risk, with a very low rate of cardiopulmonary comorbidities (<1%) reflecting an American Society of Anesthesiology (ASA) score ≤II in 99.9% of patients. Mean LD age was 31 (standard deviation [SD] 10) years, with 61% males. The mean body mass index (BMI) was 23.5 kg/m² [3.5] reflecting a stringent selection policy. Preoperative laboratory values were within normal range in all patients, showing for example a mean bilirubin level of 0.8 mg/dL [0.4] (13.6 μmol/L [6.8]) and mean INR of 1.0. Mean hematocrit and creatinine levels were 42% [4.2] and 0.8 mg/dL [0.17] (70.4 μmol/L [15]), respectively.

Characteristics of Hepatectomy

The characteristics of surgical procedures are summarized in Table 2. Procedures were performed through laparotomy and fulfilled the criteria for anatomic RH or LH. Mean duration of surgery was 7.7 [2] hours (7.5 [2] in RH vs 8.3 [3] in LH). The modified Makuuchi incision, also labeled as reverse L-incision—that is, without thoracotomy—was the most frequently used access (71% in RH vs

TABLE 2. Surgery-related Characteristics

	All (n = 5202)	RH (n = 4206)	LH (n = 996)
Type of incision*			
Modified Makuuchi†	3734 [71.8%]	2983 [70.9%]	751 [75.4%]
Subcostal	810 [15.6%]	787 [18.7%]	23 [2.3%]
Midline	55 [1.1%]	22 [0.5%]	33 [3.3%]
Mercedes	45 [0.9%]	40 [1.0%]	5 [0.5%]
Pringle maneuver (%)	147 [2.8%]	100 [2.4%]	47 [4.7%]
Mid hepatic vein taken with the graft (%)	1169 [22.5%]	239 [5.7%]	930 [93.4%]
Transsection device‡			
Cusa	4483 [86.2%]	3513 [83.5%]	970 [97.4%]
Hydrojet	586 [11.3%]	568 [13.5%]	18 [1.8%]
Kelly-crush clamp	5 [0.1%]	4 [0.1%]	1 [0.1%]
Surgery time, min	461 [122]	451 [102]	501 [176]
Intraoperative transfusions (%)	109 [2.1%]	100 [2.4%]	9 [0.9%]

*Unclear information or combined in 10.6%.

†Also labeled as reverse-L incision; does not include thoracotomy.

‡Unclear information in 2.4%.

Values in square brackets are SDs.

75% in LH), almost exclusively used in Asia (99%), followed by the subcostal incision (19% in RH vs 4.4% in LH), the preferred incision in Europe and North America (71%). Cavitron Ultrasonic Surgical Aspirator was the most commonly used transection device (83% in RH vs 97% in LH), followed by hydrojet (11%, $n = 586$). One hundred nine LD (2.1%) received intraoperative blood transfusions with lower rate of transfusion in LH (0.9%) versus 2.4% in RH. Pringle maneuver was used in 2.4% of RH and 4.7% of LH mostly for single short periods. Mid hepatic vein was taken in 5.7% of RH and 93.4% of LH. No hepatic vein reconstructions have been performed in donors.

Morbidity After RH and LH

Mean length of hospital stay was 11.7 [5.3] days (11.7 [5.6] in RH vs 11.8 [4.1] in LH). Transfusion after surgery was a rare event with only 28 patients (0.5%) requiring blood (0.6% in RH vs 0.1% in LH), and only 2 receiving more than 2 erythrocyte concentrates. Outcome after RH and LH is summarized in supplementary Table 6, <http://links.lww.com/SLA/B46>. Minor complications (\leq grade II) were comparable after RH and LH, but major complications (\geq grade III) were significantly higher after RH. Incidence rate of major complications after RH at discharge, 3 and 6 months, was 3%, 4%, 4.1%, respectively, versus 1.4% ($P = 0.001$), 2.0% ($P = 0.008$), and 2.2% ($P = 0.014$) after LH.

Seventy five patients (1.5%) developed a complication \geq IIIb, most occurred during hospitalization ($n = 59$, 1.1%), whereas 18 occurred between discharge and the third postoperative month, and another 7 between 3 and 6 months. The incidence rate for complications \geq IIIb after RH and LH at discharge, 3 and 6 months was 1.3%, 1.4%, 1.5% versus 0.6%, 0.8%, 1.0%, respectively ($P = 0.03$, $P = 0.14$, $P = 0.3$).

Complications grade \geq IV ($n = 7$, 0.13%) were only observed after RH, which included cardiopulmonary complications ($n = 4$), inferior vena cava thrombosis ($n = 2$), and 1 death (grade V) related to intraoperative cardiac arrest. No other fatal outcome occurred in the follow-up showing an overall mortality of 0.02% in the 3 or 6 months follow-up.

The mean CCI after RH and in patients who developed a complication at discharge, 3 and 6 months was 20.6, 21.6 and 21.8, respectively. In contrast after LH, those figures were 18.5 ($P = 0.17$, when compared to RH), 19.1 ($P = 0.05$) and 19.4 ($P = 0.08$).

Incidence and Severity of Complications

During the follow-up, 628 LD (12%) developed 817 postoperative complications; 13% ($n = 536$) after RH versus 9.2% ($n = 92$) after LH ($P < 0.01$). Although 488 LD (417 (9.9%) after RH and 71 (7.1%) after LH developed complications during the initial hospitalization, 140 patients developed their complications after initial discharge, mostly due to cardiopulmonary and biliary events, as well as wound infections. Between 3 and 6 months after surgery, only 6 operative revisions were required, for incisional hernias ($n = 4$), re-laparotomy for adhesive small bowel obstruction ($n = 1$) or for bile duct stenosis ($n = 1$).

As shown in supplementary Table 7, <http://links.lww.com/SLA/B46>, the most common types of complications were of biliary ($n = 146$, 18% of total complications) and cardiopulmonary ($n = 115$, 14%) origins. Considering the total number of 248 major complications, biliary ($n = 79$, 32%) and cardiopulmonary ($n = 47$, 19%) events as well as abdominal bleeding ($n = 31$, 12.5%) and intra-abdominal collections ($n = 31$, 12.5%) represented more than two-thirds of all major complications. Minor complications were predominantly related to biliary problems ($n = 65$, 11.5%) or wound infections ($n = 86$, 15%).

Looking at biliary complications, 87% ($n = 127$) occurred after RH vs 13% ($n = 19$) after LH. 44.5% ($n = 65$) of biliary complications were minor and did not require operative management. Biliary drainage or stenting (grade IIIa) was performed in 29 patients after RH, whereas 16 patients required re-operations (grade IIIb). Four LD developed biliary strictures, which were treated conservatively ($n = 1$), by biliary stenting ($n = 1$) and reoperations after 2 and 14 weeks after LH ($n = 2$). From discharge to 6 months follow-up, 37 patients developed 43 biliary complications, only 2 of them after LH. Most of them were grade IIIa due to biliary leakage or bilioma. Two LD's required reoperation due to biliary strictures after discharge.

Regarding the 115 patients (2.2%), who developed cardiopulmonary complications, the vast majority occurred after RH ($n = 110$; 96%), whereas only in 4% ($n = 5$) after LH. Most of them required thoracocentesis (grade IIIa), mostly (>90%) after RH.

Looking at the 75 patients (1.4%), who developed more severe events (\geq grade IIIb), 27 were re-laparotomies due to postoperative bleeding; 17 due to biliary leakage, bile duct injuries or strictures. One patient required re-laparotomy due to ruptured biliary drain and 2 donors required T-tube placement for injuries of the common bile duct.

Influence of Center Volume on Donor Outcome

Significant association was observed between center volume and number of complications. Centers having performed ≥ 100 hepatectomies had significantly lower rates for overall (10.2% vs 35.9%, $P < 0.001$) and major complications (3% vs 12.1%, $P < 0.001$). Furthermore overall CCI and CCI in patients with complications was significantly lower in high-volume centers (2.1 vs 8.5 and 20.7 vs 23.6 with $P < 0.001$ and $P < 0.003$, respectively).

Post-hepatectomy Liver Failure

Reflecting the overall healthy population and low complication rate, PHLF was a rare event and occurred almost exclusively after RH with discrepancies in incidence according to the 4 definitions evaluated, ISGLS, 50/50, bilirubin 7 and our criteria for SFSS (Table 3). No patient experienced PHLF according to the "50–50" criteria. Ten patients with RH (0.2%) met the day 5 bilirubin >7 criteria, only 1 of them had associated complications according to the Clavien-Dindo grading system in the form of a paracentesis (grade IIIa) for ascites. Two patients recovered spontaneously from the jaundice within 2 weeks and 2 patients underwent an endoscopic retrograde cholangiography, which failed to show any abnormalities. One patient met our criteria for SFSS (0.02%), with bilirubin levels and INR above the limit on 3 consecutive days and required paracentesis (grade IIIa) for ascites.

According to the ISGLS criteria, 259 LD (5%) developed PHLF with the vast majority after RH (99%, $n = 256$) vs only 1% ($n = 3$) after LH; $P < 0.01$). Three-quarters of these cases ($n = 204$, 79%) did not require changes in post-operative management, that is, not captured by the complication grading system, whereas 55 patients (21%) developed complications within 3 months, most of them cardio-pulmonary or urinary tract complications, not obviously related to liver failure. Thus, these 259 patients fulfilling the ISGLS criteria developed 21 (27.5%) grade I, 28 (37%) grade II, 19 (25%) grade IIIa, and 8 grade IIIb (10.5%) complications.

Benchmark value for patients positive according to the ISGLS criteria is 4.8% and 3.4% for those associated with at least 1 complication according to Clavien-Dindo grading system.

Postoperative Laboratory Profile

Laboratory values are summarized in Table 3. Patients after RH showed significant higher aspartate aminotransferase and

TABLE 3. Postoperative Laboratory Results and PHLF

	RH	LH	P
Peak aspartate aminotransferase (reference <35 U/L)	207 [149]	204 [156]	<0.001
Peak alanine aminotransferase (reference <35 U/L)	222 [726]	234 [187]	0.37
Peak bilirubin (reference <21 μ mol/L/<1.2 mg/dL)	49.3 [28.9] / 2.9 [1.7]	30.6 [17] / 1.8 [1]	<0.001
Peak creatinine within 48 hours after surgery (reference <80 μ mol/L/0.9 mg/dL)	63.7 [19.5] / 0.72 [0.22]	63.7 [15] / 0.72 [0.17]	0.026
Creatinine after 1 week (reference <80 μ mol/L/0.9 mg/dL)	64.5 [18.5] / 0.73 [0.21]	67.2 [16] / 0.76 [0.18]	0.020
Peak INR	1.5 [1.1]	1.5 [5.7]	0.88
Nadir platelets (G/L)	154 [39]	170 [38]	<0.001
4 definitions of liver failure:			
ISGLS criteria, n (%)	256 (6)	3 (0.3)	<0.01
50–50 criteria, n (%)	0	0	
Bilirubin day 5 >7 criteria, n (%)	10 (0.24)	0	
SFSS, n (%)	1 (0.02)	0	

All results reported as mean with SD in square brackets.

ISGLS criteria for PHLF: When INR and Bilirubin are above the normal cut-off as defined by local laboratory on day 5 after liver resection. Due to many different laboratory standards, we set INR ≥ 1.3 and bilirubin ≥ 1.2 mg/dL.

50–50–criteria are considered positive with INR ≥ 1.7 and serum bilirubin ≥ 2.9 mg/dL on postoperative day 5. Bilirubin >7 mg/dL on the 5th postoperative day: positive, when bilirubin >7 mg/dL in noncirrhotic, noncholestatic patients.

Criteria for SFSS: positive, with 2 of the following criteria on 3 consecutive days: bilirubin >5.9 mg/dL, INR >2, encephalopathy grade 3 or 4.

bilirubin levels ($P < 0.001$), as well as lower mean platelets nadir ($P < 0.001$). Mean lowest INR was 1.5 in both groups. Mean creatinine levels peaked within 48 hours after surgery and values were comparable between RH and LH groups. Creatinine levels normalized in >99% of patients within 1 week of surgery.

Hospital Readmission

Readmission rate within 30 days was 2.8% ($n = 134$), with 2.6% after RH ($n = 111$) and 2.3% after LH ($n = 23$) ($P = 0.09$). Main causes for readmission within 30 days were biliary or cardiopulmonary problems. The most common presentation at readmission was fever and abdominal pain. Of all readmitted patients, 11 (8.2%) required reoperation at readmission. Three patients underwent incisional hernia repair and 3 cases of adhesiolysis due to bowel obstruction. Two patients were re-operated for biliary leakage with peritonitis and 2 for biliary strictures at 2 and 14 weeks after operation. One patient required major intervention due to deep vein thrombosis.

Benchmark Values

Benchmark values target postoperative morbidity including complications by grade of severity, minor (\leq II) and major (\geq IIIa) complications, as well as the CCI in patients having developed at least 1 complication (Table 4). The upper value for overall complication rate is 31%, minor complications 23%, and major complications 8.1%. At the time of discharge, these values were,

respectively, 27%, 19%, and 6%, with then only negligible changes between 3 and 6 months. Benchmarks for CCI in patients with complications at discharge, 3 and 6 months, were 27.9, 32.6, and 32.7, respectively.

DISCUSSION

Based on this multicentric study, we propose a new concept for benchmarking the best available outcome for complex surgical procedures serving as reference values for comparisons. To benchmark major liver resection, we chose anatomic right and left hemihepatectomies, performed under the optimal conditions of living donation.^{11–16} The main findings were a low morbidity rate with only 12% of patients developing a complication, a low CCI indicating most minor events, with more serious complications after RH compared with LH.

To secure credible data collection, we enrolled a large number of patients and selected centers for excellence from all around the world. Although the criteria for “excellence” remain vague and subjective, we agreed on having high volume liver centers that have experience with at least 50 LD operations and hold prospective databases, from which all relevant information could be gathered. Another condition was previous publications critically reporting on the outcome of liver surgery.

Many stakeholders in medicine, including patients, are requesting data on quality. This has resulted in mandatory reporting in national database, audits, and a rise in self-reporting in the lay and medical literature. In a number of countries, data are made publically available, and the respective centers—or sometimes surgeons—are compared with the average or median values covering the whole country. Usually a single parameter “1-year mortality” or “perioperative mortality” is reported. Although such policy enables comparisons of individual centers with the country benchmark, the obvious drawbacks are at least threefold: first the lack of credible comparison with larger number of patients and extranational centers, second the reporting of only 1 single endpoint (such as mortality), and third, the absence of risk adjustment.

The novel concept of benchmarking is different, as it allows comparison with the “best possible” results for the respective procedures targeting several clinically relevant endpoints. We also incorporated the CCI, the novel metric for morbidity, integrating all complications on a scale from 0 (no complication) to 100 (death).^{5,6,17} The concept for such benchmarking for the “best

TABLE 4. Benchmark Values

	At Discharge	After 3 Months	After 6 Months
Any complication	26.9%	31.2%	31.2%
Major complications (\geq IIIa)	6.0%	8.1%	9.2%
Minor complications (\leq II)	18.9%	22.6%	22.6%
CCI	27.9	32.6	32.7
Liver failure			
ISGLS	4.8%	Same	—
With complications	3.4%	Same	—

Values are the 75th percentile of median proportions. ISGLS, defined as positive according to the International Study Group for Liver Surgery, as INR ≥ 1.3 and bilirubin ≥ 1.2 mg/dL on postoperative day 5.

outcome” is easy to understand and apply, and therefore likely to get wide acceptance.

The CCI is an important new tool to better assess the actual burden of complications on the whole postoperative course. Recording complications individually and reporting only the most severe one, as usually done in most studies⁶ fails to provide information about the cumulative (overall) morbidity. The CCI informed on all complications balancing their weight by severity. This sensitive and clinically relevant endpoint seems key for future benchmark projects, also enabling longitudinal assessment of the cumulative morbidity, for example, at discharge versus 3 or 6 months. For example, in this study, the CCI was the most sensitive marker highlighting better outcome in larger programs. CCI will even gain increasing value in analyzing procedures with higher rates of complications.⁶

Importantly, benchmarking analyses should present ranges, above which comparative results are significantly poorer, or significantly better if they are below. In the current study, involving a patient population associated with low morbidity, only the upper threshold of the range is of value to serve as the marker for acceptable versus higher complication rates in analyzing other populations. It is crucial for benchmarking analysis to select the relevant endpoints for the specific procedures which best measure the performance of a procedure. For example, benchmark endpoints for liver surgery must not only include complications and CCI but also postoperative liver failure and readmission rates. In Table 5, looking for some standardization, we propose the minimum endpoints to include in future benchmarking studies. The duration of the follow-up has also been controversial.¹⁸ It is obvious that limiting the observation to the hospital stay is insufficient due to major center variations, and a consensus seems to arise proposing a minimum of 3 months.^{19–27} In the current study, a number of events occurred between the time of discharge and 3 months postoperatively, but very few afterward corroborating the standardized use of benchmarking at 3 months.

Morbidity rate was consistent with many studies,^{14,28–32} but some showed values above the benchmark.^{33,34} For example, our morbidity rate was lower than the 38% published by the adult-to-adult living donor liver transplantation (A2ALL) group.³⁴ One possible explanation for this observation might be related to the larger number of LD (5202 vs 392). Another reason could be the lower proportion of obese donors with a lower average BMI compared with the 60% of overweight donors in the A2ALL group.³⁴ In addition, there was a lack of clearly defined standard outcome measures, as complications were only assessed as “peri- and postoperative” events, not focusing on 3 and 6 months morbidity, as now widely proposed.^{19–27} Moreover, A2ALL study group used an earlier version of complication grading, and not the most recent and widely accepted version of the Clavien-Dindo classification.³

Of course, interpretation of any outcome data must be risk-adjusted accounting for cases that mix across different populations

(eg, different centers) and time (eg, within a center). Approaches for risk adjustment need to be developed, which include appropriate patient characteristics as well as statistical models as it has been done for other fields.^{35–37} In liver surgery we expect poorer results than the benchmark values. For example, major hepatectomies for cancer, after chemotherapy, or in the presence of underlying liver diseases will be associated with higher complication rates and thereby higher CCI. In fact, the best results of specific diseases, for example, separating malignant from benign diseases or in cirrhotic patients, should be made available through another benchmarking study. Compared with other organs, benchmark values like those obtained in LD are not available because no surgery is performed in patients without diseases. Benchmarking will have to be deducted from selected high-quality programs for selective procedures.

Several additional questions can be addressed with benchmarking data. There are accumulating studies suggesting a correlation between program volume and outcome.^{38,39} Although the current study was not designed to look at the impact of center volume on outcome, we observed a significant influence of the number of cases per center and outcomes for this highly standardized procedure and low-risk population. For example, arbitrarily selecting a cutoff of 100 cases, we found significant decrease in overall and major complications as well as CCI in larger programs. Highlighting that center effect goes well beyond the selection of patients. The analyses also revealed that the rate of major complications was about twice higher in RH, when compared with LH, whereas no other differences could be detected. This has important implication in living donation favoring left donation as far as this would have negative impact on the recipient.^{12,28,34,40–43}

In liver surgery, an important endpoint—often imbedded in the reporting of complications—is postoperative liver failure, also labeled “small-for-size” syndrome,^{10,44,45} although no consensus exists on this definition. We tested 4 commonly used definitions,^{7–10} and found that only 1, ISGLS,⁷ captured some postoperative liver dysfunction in this series. The incidence was, however, low (<5%),^{46,47} and about two-thirds of them had no clinical relevance due to an uneventful course without any associated complication; corresponding to the subgroup of ISGLS A of the 3 categories (A, B and C) described in the original article.⁷ Looking at the other third of liver failure, almost half of the developed complications were unrelated to the liver failure, such as urinary tract infections. Thus, only 1% of the total population developed liver dysfunction associated with clinically relevant complications.⁴⁸ We suggest, however, that postoperative liver failure remains included in the reporting of major hepatectomy in future studies.⁴⁹ The weight on the morbidity associated with liver failure should, however, be included into the standardized description of complications, as well as CCI, to secure consistency in reporting clinically relevant information.

In this report, we also evaluated additional parameters, which might not be of major clinical relevance such as postoperative laboratory tests. For example, the observation that serum transaminases are higher after LH than RH is of little interest and the use for comparisons with future studies is questionable. We would suggest making such reporting facultative.

This study has some inherent limitations particularly regarding missing data collections. Despite the availability of a prospective database in each center and the retrospective analysis of each case with unclear information, one cannot exclude that some events were not captured.⁵⁰ The low rate of minor complications in one of the largest centers triggered a careful reevaluation of the charts, but this did not lead to significant changes in the data. Another limitation may be that the severity of complications may have been misclassified in some instances because information on the therapeutic consequences of complications were not always fully available.

TABLE 5. Proposal for Standardized Reporting for Benchmarking

1. Include large volume centers from at least 3 continents
2. Availability of a prospective database in each center
3. Standardized complication grading system by severity
4. Standardized metric for morbidity (eg, CCI)
5. Types of complications by severity
6. Assessment of postoperative organ failure (when appropriate)
7. Three months follow-up after surgery
8. Readmission rate
9. Benchmark ranges with lower and upper limits
10. Long-term oncologic results, when appropriate

In conclusion, we consider this project as an inaugural study introducing a novel concept of benchmarking in surgery. The availability of “best achievable” results should serve to conclusively benchmark the results in individual centers, over time or even with alternative nonsurgical approaches. In every analysis comparing outcomes across centers and time, risk adjustment must be used to account for different case mix and allow for proper interpretation. Future work in liver surgery should target major hepatectomies for benign versus malignant diseases and in special populations such as in cirrhotic patients for liver diseases. This benchmarking concept may now be developed for other major procedures in general surgery and in other fields.

REFERENCES

- Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: a method for continuous quality improvement in health. *Healthc Policy*. 2012;7:101–119.
- Kay JF. Health care benchmarking. *The Hong Kong Medical Diary*. 2007;12:22–27.
- Dindo D, Demartines N, Clavien PA. Classification of surgical complications. A new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004;240:205–213.
- Clavien PA, Barkun J, de Oliveira ML, et al. The Clavien-Dindo classification of surgical complications: five-year experience. *Ann Surg*. 2009;250:187–196.
- Slankamenac K, Graf R, Barkun J, et al. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. *Ann Surg*. 2013;258:1–7.
- Slankamenac K, Nederlof N, Pessaux P, et al. The comprehensive complication index: a novel and more sensitive endpoint for assessing outcome and reducing sample size in randomized controlled trials. *Ann Surg*. 2014;260:757–762.
- Rahbari NN, Garden OJ, Padbury R, et al. Posthepatectomy liver failure: a definition and grading by the International Study Group of Liver Surgery (ISGLS). *Surgery*. 2011;149:713–724.
- Balzan S, Belghiti J, Farges O, et al. The “50-50 criteria” on postoperative day 5: an accurate predictor of liver failure and death after hepatectomy. *Ann Surg*. 2005;242:824–828.
- Mullen JT, Ribero D, Reddy SK, et al. Hepatic insufficiency and mortality in 1,059 noncirrhotic patients undergoing major hepatectomy. *J Am Coll Surg*. 2007;204:854–862.
- Dahm F, Georgiev P, Clavien PA. Small-for-size syndrome after partial liver transplantation: definition, mechanisms of disease and clinical implications. *Am J Transplant*. 2005;5:2605–2610.
- Berg CL, Gillespie BW, Merion RM, et al. A2ALL Study Group. Improvement in survival associated with adult-to-adult living donor liver transplantation. *Gastroenterology*. 2007;133:1806–1813.
- Middleton PF, Duffield M, Lynch SV, et al. Living donor liver transplantation—adult donor outcomes: a systematic review. *Liver Transpl*. 2006;12:24–30.
- Lee SG. A complete treatment of adult living donor liver transplantation: a review of surgical technique and current challenges to expand indication of patients. *Am J Transplant*. 2015;15:17–38.
- Hwang S, Lee SG, Lee YJ, et al. Lessons learned from 1,000 living donor liver transplantations in a single center: how to make living donations safe. *Liver Transpl*. 2006;12:920–927.
- Manas D, Burnapp L, Andrews PA. Summary of the British Transplantation Society UK guidelines for living donor liver transplantation. *Transplantation*. 2016;100:1184–1190.
- Levy GA, Selzner N, Grant DR. Fostering liver living donor liver transplantation. *Curr Opin Organ Transplant*. 2016;21:224–230.
- Brennan MF. Postoperative complication reporting: more than mortality and morbidity. *Ann Surg*. 2013;258:8–9.
- Sotiropoulos GC, Radtke A, Molmenti EP, et al. Long-term follow-up after right hepatectomy for adult living donation and attitudes toward the procedure. *Ann Surg*. 2011;254:694–700.
- Wakabayashi G, Cherqui D, Geller DA, et al. Recommendations for laparoscopic liver resection: a report from the second international consensus conference held in Morioka. *Ann Surg*. 2015;261:619–629.
- Talsma AK, Lingsma HF, Steyerberg EW, et al. The 30-day versus in-hospital and 90-day mortality after esophagectomy as indicators for quality of care. *Ann Surg*. 2014;260:267–273.
- Walters DM, McMurry TL, Isbell JM, et al. Understanding mortality as a quality indicator after esophagectomy. *Ann Thorac Surg*. 2014;98:506–511.
- Pezzi CM, Mallin K, Mendez AS, et al. Ninety-day mortality after resection for lung cancer is nearly double 30-day mortality. *J Thorac Cardiovasc Surg*. 2014;148:2269–2277.
- Swanson RS, Pezzi CM, Mallin K, et al. The 90-day mortality after pancreatectomy for cancer is double the 30-day mortality: more than 20,000 resections from the national cancer data base. *Ann Surg Oncol*. 2014;21:4059–4067.
- Schadde E, Raptis DA, Schnitzbauer AA, et al. Prediction of mortality after ALPPS stage-I: an analysis of 320 patients from the International ALPPS Registry. *Ann Surg*. 2015;262:780–785.
- Schiergens TS, Dörsch M, Mittermeier L, et al. Thirty-day mortality leads to underestimation of postoperative death after liver resection: a novel method to define the acute postoperative period. *Surgery*. 2015;158:1530–1537.
- Egger ME, Ohlendorf JM, Scoggins CR, et al. Assessment of the reporting of quality and outcome measures in hepatic resections: a call for 90-day reporting in all hepatectomy series. *HPB (Oxford)*. 2015;17:839–845.
- Mise Y, Vauthey JN, Zimmiti G, et al. Ninety-day postoperative mortality is a legitimate measure of hepatopancreatobiliary surgical quality. *Ann Surg*. 2015;262:1071–1078.
- Hashikura Y, Ichida T, Umeshita K, et al. Donor complications associated with living donor liver transplantation in Japan. *Transplantation*. 2009;88:110–114.
- Moon DB, Lee SG, Hwang S, et al. More than 300 consecutive living donor liver transplants a year at a single center. *Transplant Proc*. 2013;45:1942–1947.
- Lo CM, Fan ST, Liu CL, et al. Lessons learned from 100 right lobe living donors liver transplants. *Ann Surg*. 2004;240:151–158.
- Muzaale AD, Dagher NN, Montgomery RA, et al. Estimates of early death, acute liver failure, and long-term mortality among live liver donors. *Gastroenterology*. 2012;142:273–280.
- Ibrahim S, Chen CL, Lin CC, et al. Intraoperative blood loss is a risk factor for complications in donors after living donor hepatectomy. *Liver Transpl*. 2006;12:950–957.
- Cheah YL, Simpson MA, Pomposelli JJ, et al. Incidence of death and potentially life-threatening near-miss events in living donor hepatic lobectomy: a world-wide survey. *Liver Transpl*. 2013;19:499–506.
- Ghobrial RM, Freise CF, Trotter JF, et al. A2ALL Study Group. Donor morbidity after living donation for liver transplantation. *Gastroenterology*. 2008;135:468–476.
- Zupancic JAF, Richardson DK, Horbar JD, et al. Revalidation of the score for neonatal acute physiology in the Vermont Oxford Network. *Pediatrics*. 2007;119:156–163.
- Rose S, Zaslavsky AM, McWilliams JM. Variation in accountable care organization spending and sensitivity to risk adjustment: implications for benchmarking. *Health Aff (Millwood)*. 2016;35:440–448.
- Paul E, Bailey M, Kasza J, et al. The ANZROD model: better benchmarking of ICU outcomes and detection of outliers. *Crit Care Resusc*. 2016;18:25–36.
- Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346:1128–1137.
- Birkmeyer JD, Stukel TA, Siewers AE, et al. Surgeon volume and operative mortality in the United States. *N Engl J Med*. 2003;349:2117–2127.
- Taketomi A, Kayashima H, Soejima Y, et al. Donor risk in adult-to-adult living donor liver transplantation: impact of left lobe graft. *Transplantation*. 2009;87:445–450.
- Kousoulas L, Becker T, Richter N, et al. Living donor liver transplantation: effect of the type of liver graft donation on donor mortality and morbidity. *Transpl Int*. 2011;24:251–258.
- Marsh JW, Gray E, Ness R. Complications of right lobe living donor liver transplantation. *J Hepatol*. 2009;51:715–724.
- Lo CM. Complications and long-term outcome of living liver donors: a survey of 1,508 cases in five Asian centers. *Transplantation*. 2003;75:12–15.
- Clavien PA, Oberkofler CE, Raptis DA, et al. What is critical for liver surgery and partial liver transplantation: size or quality? *Hepatology*. 2010;52:715–729.
- Clavien PA, Petrowsky H, de Oliveira ML, et al. Strategies for safer liver surgery and partial liver transplantation. *N Engl J Med*. 2007;356:1545–1559.

46. Kuramitsu K, Fukumoto T, Kobayashi T, et al. The incidence of posthepatectomy liver failure defined by the international study group of liver surgery among living donors. *J Gastrointest Surg.* 2016;20:757–764.
47. Schnitzbauer AA, Mönch C, Meister G, et al. Incidence of posthepatectomy liver failure and biliary leakage: a cohort study. *Chirurg.* 2015;86:776–780.
48. Skrzypeczek C, Truant S, Duhamel A, et al. Relevance of the ISGLS definition of posthepatectomy liver failure in early prediction of poor outcome after liver resection: study on 680 hepatectomies. *Ann Surg.* 2014;260:865–870.
49. Lafaro K, Buettner S, Magsood H, et al. Defining post hepatectomy liver insufficiency: where do we stand? *J Gastrointest Surg.* 2015;19:2079–2092.
50. Clavien PA, Puhon MA. Biased reporting in surgery. *Br J Surg.* 2014;101:591–592.

DISCUSSANTS

W.C. Chapman (St. Louis, MO):

I have 2 disclosures. I am one of several founders of an image guidance technology, Pathfinder Therapeutics, and I am on an advisory board for Novartis Pharmaceuticals.

I very much appreciate the privilege of discussing this important article on benchmarking for complications after LD hepatectomy in over 5000 patients from 12 high-volume centers. As we have heard, the authors assessed endpoints out to 6 months postoperatively assessing morbidity defined by the Clavien-Dindo classification, as well as new comprehensive classification index, or CCI, and liver failure according to several different definitions.

Strikingly, the authors found an impressively low rate of complications at 12% overall with centers having performed more than 100 hepatectomies, having significantly lower rates of overall complications, and major complications, as well as a significant reduction in the overall CCI. The authors conclude that this outcome analysis of healthy LDs may serve as a reference for performance in patients undergoing major liver resection, including across different patient populations and centers.

I have several questions for the authors.

First, how were the thresholds for individual complications defined, and who gathered the complications in individual centers? It is noted there was significant variability in complications reported by center.

There has been a significant expansion of the National Surgical Quality Improvement Project, or NSQIP, in the United States and elsewhere, and this involves the use of independent reviewers of patient outcomes for selection of complications at a predefined threshold. Is it possible that different review systems in this analysis accounted for the variability and definition of complications?

Second, the authors suggest that benchmarking will allow for comparisons to hepatectomy in other settings. As noted by the authors, this is a very different patient group compared with the typical patient undergoing hemihepatectomy for malignancy, for example, where the patient age, liver quality, BMI, and other parameters would be expected to be quite different.

Why pick the 75th percentile as a benchmark—and you touched on this, and maybe you can touch on this just a little further—instead of the mean or median? Would this mean that 75% of programs are below the standard? And would this apply to payment or penalty purposes?

I wonder about the proposal to use this cohort as a benchmark for hepatectomy in the setting of intrinsic disease. In very few settings do we ever operate on healthy individuals who otherwise have no indication for intervention.

Next, how do these results compare with prior studies on benchmarking in pancreaticoduodenectomy, for example? There have been several previous reports on benchmarking in

comprehensive assessment. Perhaps you could comment on the comparisons of CCI to other benchmarking methods.

Finally, in relation to the results seen here today, there is a significant difference in what is described as a high-volume versus what I would say is the super-high-volume centers with the cut point of greater than 100 LD hepatectomies. Why has LD transplantation taken off at such a substantially greater level such that some centers in eastern countries, for example, have such high volume compared with the west? Is it simply a matter of availability, or lack thereof, of cadaveric organs, or are there other factors involved?

My congratulations to the authors on their excellent manuscript and to Professor Clavien for his induction into the American Surgical Association as a honorary member.

Response From P.-A. Clavien (Zurich, Switzerland):

Dr. Chapman, thank you very much for your insight into our work and your excellent questions. I will try to answer your questions in order.

The first question deals with which thresholds we used to define individual complications as well as the accuracy of data collection and independence of the data collectors at each of the 12 participating centers. Those are indeed important issues to secure complete and correct data collection. The threshold defining a complication is based on the principle of our classification of complications, that is, any deviation from the expected postoperative course is at least a grade I complication. Among the criteria to select the “so-called” centers of excellence was the availability of a prospectively collected database, although collection was not performed by independent personal at all centers. To minimize bias, the Zurich group—which performed the benchmark analysis—requested re-analysis of each patient, who developed complications. Another factor minimizing underreporting is the use of a grading system ranking complications by severity based on the treatment used to correct the negative events. Such information is readily available in most databases. I will comment below on the calculation of benchmark values.

Regarding the reference to NSQIP, there is no question that this comprehensive administrative data collection, supported by the American College of Surgeons, is a great step in outcome research with more than 500 hospitals using it in the United States and elsewhere. NSQIP, however, is complex, costly, and requires extra personal. It does not allow to assess the complication by severity or to calculate a global index for complications. Complications are recorded by diagnosis with a limited follow-up of 30 days. In the current study, we found that about 20% of complications occurred between discharge and 3 months after surgery. Considering that NSQIP provides different information, including the CCI might be of great value.

Your second question deals with risk adjustment. The purpose of this study was benchmarking for the best outcome for a procedure. This is why we chose the lower-risk population of living liver donors. It is a first step that must be followed by benchmarking of surgery in other populations such as those with malignant tumors or cirrhosis. The goal is to enable individual centers to compare their results with the best benchmarking values of the procedure, as well as risk-adjusted benchmark values.

Your question about the 75th percentile as a benchmark requires some clarifications. For the calculation of the benchmark thresholds for complications by severity and the CCI, we used the 75th percentile of the median results across centers. This means that 75% of centers had a median performance that is at or better than the benchmark, and 25%, that is, those above the benchmark perform worse than the benchmark. We chose this approach to exclude the outliers, and all centers, regardless of the size, contributed equally to the benchmark values. This method gives a somewhat better picture

of the real world and may buffer the possible underreporting of low-grade complications.

Your next question is about comparison between CCI and other benchmarking methods. I must admit that I have not read every article on benchmarking. Clearly, other studies have selected other endpoints, and usually only 1 marker such as 1- or 5-year survival, and benchmarks are mostly used for comparison of individual centers with others in the same area. I may just comment on a study we did about 10 years ago with the team of John Cameron from Johns Hopkins. At that time, we wanted to test our new classification system in their database covering 600 cases of duodenopancreatectomy performed during a 2-year period. This study presented at the ESA, enabled to weigh each type of complications by severity and have an understanding of the respective postoperative events. This was a sort of benchmarking of a single center, but we did not have the CCI at that time.

Your next question is about volume and outcome. This was not the aim of this study, but the observation that expert centers performing more versus less than hundreds cases had significantly better outcome in terms of CCI and major complications may highlight some continuity in quality with volume, well over with what is currently reported. Your last point, which is not really related to the current benchmark analyses, is why LD is more popular in Asia than it is in other parts of the world? I can only speculate on that: the main reason is the lack of available cadaveric organs mostly due to cultural and religious beliefs and high mortality on the waiting list.

J. Emond (New York, NY):

This work was beautifully presented, as always, and very ambitious. I would express my ambivalence about the ability to apply a benchmark to a population with such center variability in terms of complication rates. On the table with the individual centers, those bars were quite disparate.

I think the benchmark may well have to be established with the big center. Comparing a center with 3000 donors, it is really hard for me to be comfortable with the choice of an equivalent of 100 as the benchmark, or I should say the transition point. So I would like to

challenge you to think about that a little further and whether another way to handle this ought to be considered.

The other thing, I was reassured to see that the left lobes had fewer problems. In our own data—and unfortunately we did not share it today—but both functional parameters are much better for LH than right, and you showed this, although the incidence of severe complications may not be so different, so that is something that your paper will cast a lot of light on and be very useful.

The last thing I would like to ask you is how do you evaluate the challenge of, as expertise increases, the surgeon chooses to take on more difficult or higher-risk procedures? That does not apply here because the donor hepatectomy is the same, but in other types of cancer surgery it would be a really interesting issue.

Response From P.-A. Clavien (Zurich, Switzerland):

Dr. Emond, thank you for your interest in our study. Regarding your first question, I reacted exactly like you when I received the data with high variation among the 12 centers, and not just between low and high volume centers. Dr. Milo Puhon, our epidemiologist–statistician of the study, convinced me that this was expected and that the methodology—as discussed in my reply to Dr. Chapman’s comments—based on the median value of the 75% percentile results among centers would provide not just generous but realistic values. The key is to upfront define criteria (as described in the last table of the article), then gather credible and robust data to calculate benchmark values. Those represent supposedly the best achievable results for the procedure in the real world.

Your last question about why some surgeons may take more risk while climbing the learning curve is somewhat intuitive. My feeling is that surgeons may refine their indications and technical skills while gaining experience with a procedure, and consequently feel more comfortable to take more difficult cases and perform extended surgery. This is possibly where benchmarking may serve as the “gatekeeper.”

Thank you so much to both discussants for their insightful questions and the privilege of the floor.